EVALUATION OF SAR PREDICTIONS OF ESTROGEN RECEPTOR BINDING AFFINITY

EPA CONTRACT NUMBER 68-W-01-023
WORK ASSIGNMENT 2-3

August 1, 2002

[chemical IDs removed from this version – jpk]

PREPARED FOR

Jim Kariya
WORK ASSIGNMENT MANAGER
U.S. ENVIRONMENTAL PROTECTION AGENCY
ENDOCRINE DISRUPTOR SCREENING PROGRAM
WASHINGTON, D.C.

BATTELLE
505 KING AVENUE
COLUMBUS, OH 43201

Evaluation of SAR Predictions of Estrogen Receptor Binding Activity
Work Assignment 2-3

I.0   Introduction and Background

The Environmental Protection Agency (EPA) is attempting to validate Structure Activity Relation (SAR) models to predict the extent of binding affinity of chemicals with estrogen receptor sites. Two models are considered in this report, referred to as "Model A" and "Model B".

The validation experiment was carried out by comparing the model predictions of Relative Binding Affinity (RBA) for a sample of chemicals with estrogen receptor (ER) binding affinity laboratory assay results. The laboratory assay results are taken to be the authority measure.

This report discusses the results of a statistical comparison of the laboratory assay results with the model predictions. Estimates of sensitivity, specificity, positive predictive probability, and negative predictive probability of each of the models are determined for various portions of the data and are compared among one another. These measures are defined and discussed in some detail in the report "Issues Related to Sampling Chemicals for Verifying Predictiveness of Endocrine Binding Activity QSAR Models", August 17, 2001.   The effects of using both models jointly for prediction of positive ER activity is discussed.

Section II discusses the sources of data underlying the comparisons and data decisions that were made concerning which portions of the data to include in the comparisons and which portions to exclude. Section III discusses the results of the detailed comparisons that were carried out. The section is divided into subsections in accordance with items 1 to 7 in the "Detailed description of statistical analyses being requested" section of EPA's QSAR model evaluation Work Assignment dated June 19, 2002. The discussion in Section III is based on and refers to the tables, figures, and detailed calculations included in Appendix A.

2.0   DATA

The EPA selected a set of 9,067 chemicals for which it needs to set priorities for testing for endocrine receptor binding activity. Model A and Model B made predictions of RBA on a subset of 6,649 chemical from this set, for which CAS numbers (and therefore chemical structure specification) exist. Each of the models divided these chemicals into six strata (differing for each model) depending on the order of magnitude of the predicted RBA. Model A predicted 319 of the 6,649 chemicals (4.8%) to be positive endocrine receptor binders. Model B predicted 304 of the 6,649 chemicals (4.6%) to be binders. These was an overlap of 78 chemicals (1.2%) between the positive predictions from each model.

Three samples of chemicals were selected from the subset of 6,649 chemicals for evaluation by laboratory assay.

The first sample was (ideally) a simple random sample of (nominally) 200 chemicals from the subset of 6,649. Actually 197 chemicals were selected. Several deviations from the random sampling scheme were necessitated by difficulties in acquiring some of the chemicals or in acquiring them at the purity required for the assay. Other chemicals from the randomly generated list were substituted until a set of nearly 200 acquirable chemicals of the desired purity was obtained. The first sample was designed to permit comparisons between the laboratory assay results and predictions from each of the models. This sample is referred to as the "Random 200" chemicals.

The second and third samples were (ideally) stratified random samples of (nominally) 50 chemicals each from among the chemicals that Model A predicted to be positive (Model A sample) and from among the chemicals that Model B predicted to be positive (Model B sample). Actually 49 chemicals were selected in the Model A sample and 43 chemicals were selected in the Model B sample. The Model A sample was designed to provide an enhanced positive predictive probability sample for Model A. The Model B sample was designed to provide an enhanced positive predictive probability sample for Model B. These samples are referred to as the "50 Model A" chemicals and the "50 Model B" chemicals respectively.

The laboratory assay classified each of the 197 + 49 + 43 sampled chemicals as "Binders" (B), "Extrapolated" (E), "Activity" (A), or "Non-Binder" (N) based on the maximum extent of displacement of the radiolabeled estradiol by the test chemical. The criteria used for this classification is discussed in the "Estrogen Receptor Binding Assay Overview Report" for EPA Work Assignment 3-04 Task 4, April 2002.

Among the "Random 200" chemicals, 25 of the 197 chemicals were classified by the laboratory assay as B or E. The remainder were classified as A or N. Among the "50 Model A" chemicals, 18 of the 49 chemicals were classified as B or E. Among the "50 Model B" chemicals, 12 of the 43 chemicals were classified as B or E.

The chemicals classified as A or N were treated as negatives for the purposes of the statistical analyses. The chemicals classified as B were treated as positives. The chemicals classified as E were treated as positives in one analysis and as negatives in another. Ideally it was desired in the second analysis to only classify those Es as positive for which the lower 95 percent confidence bound on percent bound at the highest test chemical concentration fell below the 50 percent level. However the analysis would be simplified if the first analysis excluded all of the Es and the second analysis included all of the Es. These are referred to as the "bookend" analyses. If there were no qualitative differences in results between the "bookend" analyses it would not be necessary to carry out the intermediate analysis.

Several chemicals were excluded from the analyses because the laboratory assays produced steep or erratic curves, as discussed in the "Estrogen Receptor Binding Assay Overview Report". Dr. Susan Laws, EPA/RTP, reviewed the laboratory assay results and specified which chemicals should be excluded from the comparisons and which should be retained. Dr. Laws' assessments are summarized in the Excel file "ER Binding Summary Data (Task 6)". The chemicals that were omitted from the analyses based on Dr. Laws' assessments and supplemented by the recommendations in the "Estrogen Receptor Binding Assay Overview Report" were omitted from both the numerators and the denominators of the calculations of model performance.

The CAS Numbers of the chemicals that were omitted are summarized below.

Table 2-1.   Battelle M Numbers and CAS Numbers of the Chemicals That Were Omitted from the Analyses

[deleted from this copy, jpk, 8-2-02]

In addition chemical   *[deleted, jpk 8-2-02]*   ("Random 200" Chemicals, non-binder) was omitted from the analysis based on the "Estrogen Receptor Binding Assay Overview Report", Section 5, where it was stated that this chemical exhibited erratic binding behavior.

Thus the analyses in this report were based on:

1. 189 chemicals from the "Random 200" chemicals
2. 48 chemicals from the "50 Model A" Positive predicted chemicals
3. 40 chemicals from the "50 Model B" positive Predicted chemicals

The CAS numbers associated with these three samples are contained in Appendix B, along with which CAS numbers were predicted to be binders or extrapolateds by the laboratory assay and which were predicted to be positives by each of the models.

Among these chemicals there were:

4. 11 binders and 7 extrapolateds from the "Random 200" chemicals
5. 16 binders and 1 extrapolated from the "50 Positive Predicted" Model A chemicals
6. 9 binders and 0 extrapolated from the "50 Positive Predicted" Model B chemicals

The CAS numbers corresponding to these chemicals are summarized in Table 2.2.

Table 2-2. CAS Numbers of the Chemicals That Were Included in the Analyses

[deleted from this copy – jpk, 8-2-02]

Model A 's predictions of positive chemicals included 6 chemicals among the "Random 200" chemicals sample. None coincided with the 11 chemicals classified by the laboratory assay as binders and 2 coincided the 7 chemicals classified by the laboratory assay as extrapolateds.

Model B 's predictions of positive chemicals included no chemicals among the "Random 200" chemicals sample.

3.0   RESULTS AND DISCUSSION

This section discusses the results of the calculations to assess the relations between the model predictions of Relative Binding Affinity with the laboratory assay results on the same chemicals. The results are divided into subsections, numbered 1 to 7. These subsections correspond to the items enumerated in the Work Assignment section "Detailed description of statistical analyses being requested".

The tables, calculations, and figures that present the detailed results are included in Appendix A. The discussion in this section refers to those exhibits. Note that the confidence bounds shown in Appendix A are upper and lower 95% bounds. Thus the confidence intervals are 90% intervals.

1., 2. Sensitivity, Specificity, Positive Predicitivity, Negative Predicitivity of Model A and Model B Proportion of True Positives. Comparison of the Positive Predictivities Estimated from the "Random 200" Chemicals and from the "50 Model A" or "50 Model B" Samples.

The detailed analysis results are displayed in Tables A-1 to A-7 of Appendix A. The tables based on the "Random 200" chemicals sample provide estimates of sensitivity, specificity, PPP, and NPP. Those based on the "50 Model A" and "50 Model B" predicted positives samples provide estimates of positive predictive probability only.

The estimated proportion of true positive endocrine disruptors based on the laboratory assay results of the "Random 200" chemicals is $J = 5.8\%$ (Table A-1) if only the binders are included and is $J = 9.5\%$ (Table A-3) if the extrapolateds are added.

The sensitivities of both models are very low. The upper 95% confidence bound on sensitivity for Model A is at most 31.0% (Table A-3) and that for Model B, is at most 23.8% (Table A-5).

The specificities of both models are in the mid to upper 90% range. The lower 95% confidence bound on specificity for Model A exceeds 93.5% (Tables A-1, A-3) and that for Model B exceeds 98.3% (Tables A-5, A-7).

The positive predictive probabilities of both models are low. For model A, the estimated positive predictive probabilities are 33.3% based on the "Random 200" chemicals (Table A-3)

and 35.4% based on the "50 Model A" predicted positives (Table A-4), even when the extrapolateds are included in the calculation. The results based on the "Random 200" chemicals and on the "50 Model A" predicted positives are not significantly different (p=0.16, p=1.00) whether or not the extrapolateds are included. For Model B the positive predictive probability is undefined based on the "Random 200" chemicals since none of the chemicals in this sample were predicted by the model to be positive. Based on the "50 Model B" predicted positive chemicals the positive predictive probability is 22.5% (Table A-6).

The negative predictive probabilities of both models are in the 90 percent range. For model A the estimated NPP is 93.9% if the extrapolateds are not included (Table A-1) and is 91.3% if the extrapolateds are included (Table A-3). This is approximately what one would expect from choosing chemicals at random, based on the estimated values of the proportion, J, of true positive endocrine disruptors. The NPP for Model B is about the same as that for Model A, 94.2% if the extrapolateds are not included (Table A-5) and 90.5% if the extrapolateds are included (Table A-7). This is again approximately what one would expect from choosing chemicals at random, based on the estimated values of the proportion, J, of true positive endocrine disruptors.

3. Quantify the Sensitivity, Specificity, Positive Predictivity, and Negative Predictivity of Joint Predictions by the Two Models

The detailed results are displayed in Tables A-8 to A-12 of the Appendix.

Meaningful joint model predictions cannot be based on the "Random 200" chemicals (Tables A-8, A-9) sample since model B has no predicted positives among that sample.

Only positive predictive probability can be estimated from the "50 Model A" and the "50 Model B" positive predictive samples. If both models are required to be positive to infer that a chemical is positive the estimated positive predicted probability is 62.5% (Tables A-11, A-12). However just 8 chemicals of the 80 positive predictive chemicals would be jointly inferred to be positive. If just one model is required to infer that a chemical is positive the estimated positive predicted probability is 26% to 27%, depending on whether or not the extrapolateds are included (Tables A-11, A-12). The upper 95% confidence bound is approximately 35%. This is not an improvement over the individual model predictions..

4. Efficiency of Model A and Model B in Concentrating the True Positives in the Predicted Positive Set and in Diluting the True Positives in the Predicted Negative Set.

Criteria for assessing the efficiency of a model are discussed in Appendix A. These are referred to as

7. Positive Prediction Concentration Efficiency / PPP/J
8. Negative Prediction Dilution Efficiency / (1-NPP)/J

One would expect PPP/J to be greater than 1 and (1-NPP)/J to be less than 1.

Based on the "Random 200" chemical sample the estimate of the Positive Prediction Concentration Efficiency is based on 6 chemicals for Model A and 0 chemicals for Model B. Thus it is undefined for Model B and has a confidence interval ranging from approximately 0 to 7 for Model A. Thus inferences about Prediction Concentration Efficiency cannot be based on this sample.

Based on the "50 Model A" and the "50 Model B" positive predictive samples the Prediction Concentration Efficiency is estimated to be 5.7 for Model A with binders only, with lower and upper confidence bounds (3.8, 8.0). If the extrapolated chemicals are included the estimates diminish a bit, but still substantially exceed 1. For Model B the Prediction Concentration Efficiency is estimated to be 3.9, with lower and upper confidence bounds (2.1, 6.2). Both models have at least moderate Positive Prediction Concentration Efficiency.

Based on the "Random 200" chemical sample the estimate of the Negative Prediction Dilution Efficiency is 1.0 or 0.9 for Model A (depending on whether extrapolateds are included) and 1.0 for Model B. Confidence bounds range from about 0.6 at the lower end to about 1.4 to 1.7 at the upper end. Thus neither model decreases the probability that a chemical is, in fact, positive, conditional on the model predicting the chemical to be negative.

5. Relationship Between Predicted Binding Strength and Positive Predictivity

The relationship is displayed in Appendix A, Tables A-13 to A-15. The relationship cannot be assessed based on the "Random 200" chemical sample because the six positive predictions by Model A all fall within the weakest stratum, $\log_{10}$(RBA) $\bigcirc$ [-3, -2]. For Model B there were no positive predictions.

For the "50 Model A" and "50 Model B" positive prediction chemical samples the trend in PPP with RBA is nonsignificant for Model A (p=1.0) (Table A-14) and marginally significant for Model B (p=0.09) (Table A-15). Note however that the trend in Model B is opposite to what one would expect. The highest positive predictive probability occurs in the weakest RBA stratum.

We thus conclude that these models do not demonstrate association between predicted binding strength and positive predictivity.

6. Degree of Overlap of Positive Predictions Between Model A and Model B

The degree of overlap between the positive predictions for Model A and Model B is displayed in Appendix A, Tables A-16 to A-18, particularly Table A-18.
.

Each model predicted about 300 of the 6,649 chemicals to be positive. There were 78 chemicals that were predicted to be positive by both models. This is approximately 25% of each model's predictions.

When dividing predictions into RBA strata Table A-18 demonstrates that Model B generally predicted greater RBA than Model A. 21 of the 78 predictions were in the same stratum; 10 of the 21 in the weakest stratum, [-3, -2]. While there were just nine chemicals for which Model A predicted a higher stratum than Model B, there were 48 chemicals for which Model B predicted a higher stratum than Model A.

There is not a great deal of stratum overlap of positive predictions between Model A and Model B.

7. Relationship Between Measured Binding Strength and Standard Error of the RBA.

The results in this section are based on the laboratory results only. Standard errors of $\log_{10}$ RBA estimates were available only for the binders and for the extrapolated chemicals. Those chemicals that were excluded from the previous analyses because of steep or erratic binding curves are also excluded from the analysis in this section.

The relationship between $\log_{10}$ RBA and standard error of $\log_{10}$ RBA is displayed in Appendix A, Figures A-1 to A-4. Figure A-1 pertains to the "200 Random" chemical sample. Figures A-2 and A-3 pertain to the "50 Model A" and "50 Model B" positive prediction chemical samples respectively. Figure A-4 displays chemicals from all three samples superimposed to assess whether there were any differences in the relationships. In Figures A-1 to A-3 the binders and extrapolateds were plotted using symbols "B" and "E" respectively. The "E"s are seen to have lower RBAs than the "B"s, as would be expected.

Correlation coefficients and associated p-values between average $\log_{10}$ RBA and average standard error $\log_{10}$ RBA are shown below in Table 2.3.

**Table 2.3 Correlation Coefficients and Associated P-values Between
Average Log$_{10}$ RBA and Average Standard Error Log$_{10}$ RBA**

"Random 200" Chemicals	"50 Positive Predicted" Model A Chemicals	"50 Positive Predicted" Model B Chemicals
0.36 (0.14) Figure A-1	-0.34 (0.18) Figure A-2	-0.52 (0.15) Figure A-3

These is no significant association between average $\log_{10}$ RBA and average standard error $\log_{10}$ RBA for any of the samples (Figures A-1 to A-3). The relationships in all three samples coincide (Figure A-4).

4.0   REFERENCES

1.   "Issues Related to Sampling Chemicals for Verifying Predictiveness of Endocrine Binding Activity QSAR Models", August 17, 2001. Battelle. Endocrine Disruptor Screening Program Work Assignment 1-5, Task 4. Contract No. 68-W-01-023.

2.   "Estrogen Receptor Binding Assay Overview Report", April, 2002. Battelle. OPPT Statistical and Technical Support for the Assessment of Toxic Substances Work Assignment 3-04 Task 6. 2001. Contract No. 68-W-99-033.

3.   "ER Binding Summary Data (Task 6)". Excel file. Communication from USEPA/OSCP/OPPTS. June, 2002.

APPENDIX A

Tables, Figures, Detailed Calculations

1., 2. Sensitivity, Specificity, Positive Predicitivity, Negative Predicitivity of Model A and Model B. Proportion of True Positives. Comparison of the Positive Predictivities Estimated from the "200" Chemicals and from the Model A or Model B Samples.

A. Model A. Binders Only
  "200" Chemicals

Table A-1

		Model A Prediction		
		Positive	Negative	Total
Lab Result	Positive	0	11	11
	Negative	6	172	178
	Total	6	183	189

Sensitivity: (0/11) = 0% Lower and Upper 95% Confidence Bounds (0%, 23.8%)
Specificity:(172/178) = 96.6% Lower and Upper 95% Confidence Bounds (93.5%, 98.5%)
PPP: (0/6) = 0% Lower and Upper 95% Confidence Bounds (0%, 39.3%)
NPP: (172/183) = 93.9% Lower and Upper 95% Confidence Bounds (90.2%, 96.6%)
J = P(True Positive): (11/189) = 5.8%
                       Lower and Upper 95% Confidence Bounds (3.3%, 9.5%)

"50" Model A Positive Prediction Chemicals

Table A-2

		Model A Prediction		
		Positive	Negative	Total
Lab Result	Positive	16	0	16
	Negative	32	0	32
	Total	48	0	48

PPP: (16/48) = 33.3% Lower and Upper 95% Confidence Bounds (22.2%, 46.1%)

Two-sided comparison between PPPs (0% and 33.3%): p=0.16, not significant.

B. <u>Model A. Binders Plus Extrapolated</u>
       <u>"200" Chemicals</u>

Table A-3

		Model A Prediction		
		Positive	Negative	Total
Lab Result	Positive	2	16	18
	Negative	4	167	171
	Total	6	183	189

Sensitivity: (2/18) = 11.1%   Lower and Upper 95% Confidence Bounds (2.0%, 31.0%)
Specificity: (167/171) = 97.7%   Lower and Upper 95% Confidence Bounds (94.7%, 99.2%)
PPP: (2/6) = 33.3%   Lower and Upper 95% Confidence Bounds (6.3%, 72.9%)
NPP: (167/183) = 91.3%   Lower and Upper 95% Confidence Bounds (87.0%, 94.4%)
J = P(True Positive): (18/189) = 9.5%
                              Lower and Upper 95% Confidence Bounds (6.2%, 13.8%)

"50" Model A Positive Prediction Chemicals

Table A-4

		Model A Prediction		
		Positive	Negative	Total
Lab Result	Positive	17	0	17
	Negative	31	0	31
	Total	48	0	48

PPP: (17/48) = 35.4%   Lower and Upper 95% Confidence Bounds (24.0%, 48.3%)

Two-sided comparison between PPPs (33.3% and 35.4%): p=1.00, not significant.

C. Model B. Binders Only
  "200" Chemicals

Table A-5

		Model B Prediction		
		Positive	Negative	Total
Lab Result	Positive	0	11	11
	Negative	0	178	178
	Total	0	189	189

Sensitivity: (0/11) = 0%　　　　　Lower and Upper 95% Confidence Bounds (0%, 23.8%)
Specificity: (178/178) = 100 %　　Lower and Upper 95% Confidence Bounds (98.3%, 100 %)
PPP: (0/0) Undefined
NPP: (178/189) = 94.2%　　　　　Lower and Upper 95% Confidence Bounds (90.6%, 96.7%)
J = P(True Positive): (11/189) = 5.8%
　　　　　　　　　　　　　　　　Lower and Upper 95% Confidence Bounds (3.3%, 9.5%)

"50" Model B Positive Prediction Chemicals

Table A-6

		Model B Prediction		
		Positive	Negative	Total
Lab Result	Positive	9	0	9
	Negative	31	0	31
	Total	40	0	40

PPP: (9 /40) = 22.5%　　　　　　Lower and Upper 95% Confidence Bounds (12.3%, 36.0%)

D. Model B. Binders Plus Extrapolated
  "200" Chemicals

Table A-7

		Model B Prediction		
		Positive	Negative	Total
Lab Result	Positive	0	18	18
	Negative	0	171	171
	Total	0	189	189

Sensitivity: (0/18) = 0%   Lower and Upper 95% Confidence Bounds (0%, 15.3%)
Specificity: (171/171) = 100 %   Lower and Upper 95% Confidence Bounds (98.3%, 100 %)
PPP: (0/0) Undefined
NPP: (171/189) = 90.5%   Lower and Upper 95% Confidence Bounds (86.2 93.8%)
J = P(True Positive): (18/189) = 9.5%
                                    Lower and Upper 95% Confidence Bounds (6.2%, 13.8%)

    "50" Model B Positive Prediction Chemicals (See Section C. There were no extrapolated Model B chemicals).

 3. Quantify the Sensitivity, Specificity, Positive Predictivity, and Negative Predicitivity of Joint Predictions by the Two Models

Rule 1. A+ and B+ implies Chemical Positive
Rule 2. A+ or B+ implies Chemical Positive

A. "200" Samples. Binders Only

Table A-8

Lab	Model				
	A+, B+	A+, B-	A-, B+	A-, B-	
Positive	0	0	0	11	11
Negative	0	6	0	172	178
Total	0	6	0	183	189

Rule 1. A+ and B+ implies Chemical Positive

Sensitivity: (0/11) = 0%   Lower and Upper 95% Confidence Bounds (0%,23.8%)
Specificity: (178/178) = 100%   Lower and Upper 95% Confidence Bounds (98.3%,100%)
PPP: (0/0) Undefined
NPP: (178/189) = 94.2%   Lower and Upper 95% Confidence Bounds (90.6%,96.7%)

Rule 2. A+ or B+ implies Chemical Positive

Sensitivity: (0/11) = 0%   Lower and Upper 95% Confidence Bounds (0%,23.8%)
Specificity: (172/178) = 96.6%   Lower and Upper 95% Confidence Bounds (93.5%,98.5%)
PPP: (0/6)= 0%   Lower and Upper 95% Confidence Bounds (0%,39.3%)
NPP: (172/183) = 94.0%   Lower and Upper 95% Confidence Bounds (90.2%,96.6%)

B. "200" Samples. Binders and Extrapolated

Table A-9

Lab	Model				
	A+, B+	A+, B-	A-, B+	A-, B-	
Positive	0	2	0	16	18
Negative	0	4	0	167	171
Total	0	6	0	183	189

Rule 1. A+ and B+ implies Chemical Positive

Sensitivity: (0/18) = 0%   Lower and Upper 95% Confidence Bounds (0%,23.8%)
Specificity: (171/171) = 100%   Lower and Upper 95% Confidence Bounds (98.3%,100%)
PPP: (0/0) Undefined
NPP: (171/189) = 90.5%   Lower and Upper 95% Confidence Bounds (86.2%, 93.6%)

Rule 2. A+ or B+ implies Chemical Positive

Sensitivity: (2/18) = 11.1%   Lower and Upper 95% Confidence Bounds (2.0%, 31.0%)
Specificity: (167/171) = 97.7%   Lower and Upper 95% Confidence Bounds (94.7%,99.2%)
PPP: (2/6)= 33.3%   Lower and Upper 95% Confidence Bounds (6.3%,72.9%)
NPP: (167/183) = 91.3%   Lower and Upper 95% Confidence Bounds (87.0%,94.4%)

C. Chemicals Predicted Positive by One or Both Models. Binders Only

Based on the performance of the models on the (sub)population of 6,649 TSCA chemicals, we calculate the following joint probabilities on Model A and Model B results, conditional on Model A being positive for a chemical or Model B being positive.

Table A-10

		Model A		
		Positive	Negative	Total
Model B	Positive	78	226	304
	Negative	241		
	Total	319		545

Therefore

$$P(A+, B+ \mid A+ \text{ or } B+) = 78/545 = 0.143 = B_{++}$$
$$P(A+, B- \mid A+ \text{ or } B+) = 241/545 = 0.442 = B_{+-}$$
$$P(A-, B+ \mid A+ \text{ or } B+) = 226/545 = \underline{0.415} = B_{-+}$$
$$1.000$$

Table A-11

		SAR Model Predictions				
		A - pos. B - pos.	A - pos. B - neg.	A - neg. B - pos.	A - neg. B - neg.	Total
Lab Result	Positive	5	11	4	0	
	Negative	3	29	28	0	
	Total	8	40	32	0	80

Note that only PPP can be estimated from this table.

Rule 1. A+ and B+ implies Chemical Positive

PPP = P(Lab+|A+, B+) = 5/8 = 62.5%
Lower and Upper 95% Confidence Bounds (28.9%, 88.9%).

Rule 2. A+ or B+ implies Chemical Positive

$PPP = P(Lab+|A+ \text{ or } B+) = P(Lab+|A+,B+)B_{++} + P(Lab+|A+,B-)B_{+-} + P(Lab+|A-,B+)B_{-+} =$

(5/8)(0.143) + (11/40)(0.442) + (4/32)(0.415) = 26.3%.

This is approximately the (right) marginal ratio 20/80 = 0.25. 95 percent confidence bounds on this "probability" are (17.2%, 34.2%). We use these as approximate confidence bounds on the positive predictive probability.

D. Chemicals Predicted Positive by One or Both Models. Binders and Extrapolated

Table A-12

		SAR Model Predictions				
		A - pos. B - pos.	A - pos. B - neg.	A - neg. B - pos.	A - neg. B - neg.	Total
Lab Result	Positive	5	12	4	0	
	Negative	3	28	28	0	
	Total	8	40	32	0	80

 Rule 1. A+ and B+ implies Chemical Positive

PPP = P(Lab+|A+, B+) = 5/8 = 62.5%
Lower and Upper 95% Confidence Bounds (28.9%, 88.9%).

Rule 2. A+ or B+ implies Chemical Positive

$PPP = P(Lab+|A+ \text{ or } B+) = P(Lab+|A+,B+)B_{++} + P(Lab+|A+,B-)B_{+-} \; P(Lab+|A-,B+)B_{-+} =$

(5/8)(0.143) + (12/40)(0.442) + (4/32)(0.415) = 27.4%.

This is approximately the (right) marginal ratio 21/80 = 0.263. 95 percent confidence bounds on this "probability" are (18.3%, 35.6%). We use these as approximate confidence bounds on the positive predictive probability.

4. Efficiency of Model A and Model B in Concentrating the True Positives in the Predicted Positive Set and in Diluting the True Positives in the Predicted Negative Set.

This section discusses the performance of each model with respect to modifying the probability that a chemical is truly positive, conditional on the model predicting that the chemical is positive or conditional on it predicting that the chemical is negative. If the model performs well it would be expected that:

- The probability that a chemical is positive conditional on the model predicting it positive should be greater than the unconditional probability that a randomly chosen chemical is positive.

- The probability that a chemical is positive, conditional on the model predicting it negative should be smaller than the unconditional probability that a randomly chosen chemical is positive.

These two criteria suggest the following measures of model efficiency.

1. P(True Positive|Model Predicts Positive)/J / PPP/J
2. P(True Positive|Model Predicts Negative)/J / (1-NPP)/J

For a model that performs well it would be expected that PPP/J>>1 and (1-NPP)/J<<1. Ideally these values would be 1/J and 0 respectively.

Based on the "200" chemicals data set J is estimated as

J = 5.8% (3.3%, 9.5%)   based on the eleven binder chemicals only
J = 9.5% (6.2%, 13.8%) based on the eighteen binder and extrapolated chemicals

For purposes of the calculations in the section these estimates of J will be regarded as approximately the true population values, without variation. They are based on a relatively large sample size, n=189 chemicals.

Model A. Positive Prediction Concentration Efficiency

1. Based on "200" Chemicals Sample

Binders Only
PPP/J = (0/6)/0.058 = 0   (0/0.058, .393/0.058) = (0, 6.78)

Binders Plus Extrapolated
PPP/J = (2/6)/0.095 = 3.51   (0.063/0.095, 0.729/0.095) = (0.663, 7.67)

2. Based on "50" Chemicals Positive Prediction Sample

Binders Only
PPP/J = (16/48)/0.058 = 5.74   (0.222/0.058, .0461/0.058) = (3.82, 7.95)

Binders Plus Extrapolated
PPP/J = (17/48)/0.095 = 3.73   (0.240/0.095, 0.483/0.095) = (2.53, 5.08)

Model A. Negative Prediction Dilution Efficiency

 Based on "200" Chemicals Sample

Binders Only
(1- NPP)/J = (1 - 172/183)/0.058 = 1.04   ((1 - 0.966)/0.058, (1-0.902)/0.058) =
                                            (0.59, 1.69)

Binders Plus Extrapolated
(1 - NPP)/J = (1 - 167/183)/0.095 = 0.92   ((1 - 0.944)/0.095, (1 - 0.870)/0.095) =
                                            (0.59, 1.37)

Model B. Positive Prediction Concentration Efficiency

1. Based on "200" Chemicals Sample

Binders Only
PPP/J = (0/0)/0.058   undefined

Binders Plus Extrapolated
PPP/J = (0/0)/0.095   undefined

2. Based on "50" Chemicals Positive Prediction Sample

Binders Only

PPP/J = (9/40)/0.058 = 3.88   (0.123/0.058, .0.360/0.058) = (2.12, 6.21)

<u>Binders Plus Extrapolated</u>
Same as binders only. There were no extrapolated chemicals in the model B positive
prediction sample.

<u>Model B. Negative Prediction Dilution Efficiency</u>

<u>Based on "200" Chemicals Sample</u>

<u>Binders Only</u>
(1- NPP)/J = (1 - 178/189)/0.058 = 1.00   ((1 - 0.967)/0.058, (1-0.906)/0.058) =
                                                 (0.57, 1.62)

<u>Binders Plus Extrapolated</u>
(1 - NPP)/J = (1 - 171/189)/0.095 = 1.00   ((1 - 0.938)/0.095, (1 - 0.862)/0.095) =
                                                  (0.65, 1.45)

5. Relationship Between Predicted Binding Strength and Positive Predictivity

<u>Model A</u>.

1. <u>Based on "200" Chemicals Sample</u>

    <u>Binders Only</u>

Table A-13

		Model A Prediction						Neg	Total
		Positive $(logR_{10}$ BA)							
		>2	(1,2]	(0,1]	(-1,0]	(-2,-1]	[-3,-2]		
Lab Result	Positive	0	0	0	0	0	0	11	11
	Negative	0	0	0	0	0	6	172	178
	Total	0	0	0	0	0	6	183	189

Just 6 chemicals within the random sample of 189 chemicals were predicted to be positive. All 6

fall within the weakest binding stratum [-3, -2]. Therefore no trend can be assessed.

The same situation occurs for binders plus extrapolated.

2. <u>Based on "50" Chemicals Positive Prediction Sample</u>

     <u>Binders Only</u>

Table A-14

		Model A Prediction							
		Positive (log$_{10}$ RBA)						Neg	Total
		>2	(1,2]	(0,1]	(-1,0]	(-2,-1]	[-3,-2]		
Lab Result	Positive	0	0	0 (0%)	2 (40%)	5 (31.3%)	9 (34.6%)	0	16
	Negative	0	0	1	3	11	17	0	32
	Total	0	0	1	5	16	26	0	48

Nearly all of the chemicals are in the two weakest strata, (-2,-1] and [-3, -2]. An exact contingency table test of homogeneity of positive predictive probabilities across strata shows no significant differences (p=1.0).

There is just one extrapolated positive chemical in the positive predicted sample. It fall in the [-3, -2] stratum. Thus the positive predictive probability in that stratum becomes 10/26 = 38.5%. The exact contingency table test of homogeneity of positive predictive probabilities across strata remains nonsignificant (p=0.9).

<u>Model B</u>

1. <u>Based on "200" Chemicals Sample</u>

No chemicals among those that Model B predicted to be positive fall among the "200" chemicals random sample. Therefore no trend can be assessed. The same situation occurs for binders plus extrapolated.

2. <u>Based on "50" Chemicals Positive Prediction Sample</u>

     <u>Binders Only</u>

Table A-15

		Model B Prediction							
		Positive ($\log_{10}$ RBA)						Neg	Total
		>2	(1,2]	(0,1]	(-1,0]	(-2,-1]	[-3,-2]		
Lab Result	Positive	0	0	2 (20%)	0 (0%)	1 (10%)	6 (46.2%)	0	9
	Negative	0	0	8	7	9	7	0	31
	Total	0	0	10	7	10	13	0	40

An exact contingency table test of homogeneity of positive predictive probabilities across strata is marginally significant (p=0.09). A Mantel-Haenszel test for trend across strata is marginally significant (p=0.10). Note however that the trend is in the opposite direction from what one would expect. The largest positive predictive probability, 46.2%, is in the weakest binding stratum [-3, -2].

There are no extrapolated positive chemicals in the positive predicted sample.

6. Degree of Overlap of Positive Predictions Between Model A and Model B

Each model predicted the relative binding affinity (RBA) for the same set of 6,649 chemicals. RBAs greater than 10^{-3} % were defined as positive.

1. The marginal distributions of the positive predictions for each model by RBA strata are as follows:

Table A-16

Stratum ($\log_{10}$RBA)	>2	(1, 2]	(0, 1]	(-1, 0]	(-2, -1]	[-3, -2]	Total
Model A	0	1	7	36	121	154	319
Model B	1	23	71	63	90	56	304

The distribution of model B's RBA predictions is shifted toward higher RBAs as compared to Model A's predictions.

2. The joint frequency of positive and negative predictions is as follows:

Table A-17

		Model B		
		Positive	Negative	Total
Model A	Positive	78	241	319
	Negative	226	6,104	6,330
	Total	304	6,345	6,649

Each model predicted approximately 300 of 6,649 chemicals (4.6%, average) to be positive. Of these positive prediction chemicals 78 (approximately 25%) were predicted to be positive by both models.

2. The breakdown of the chemicals that were predicted to be positive by each model into RBA strata is as follows:

Table A-18

		Model B Prediction					
	Log$_{10}$ RBA Strata	(1, 2]	(0, 1]	(-1, 0]	(-2, -1]	[-3, -2]	Total
Model A Prediction	(0, 1]	1	**0**	1	0	0	2
	(-1, 0]	1	2	**6**	1	1	11
	(-2, -1]	4	14	7	**5**	6	36
	[-3, -2]	2	6	3	8	**10**	29
	Total	8	22	17	14	17	78

21 of the 78 jointly positive predictions overlap strata, 10 of the 21 being in the [-3,-2] stratum. For 48 of the 78 overlap chemicals Model B predicts a higher stratum than Model A.

7. Relationship Between Measured Binding Strength and Standard Error of the RBA

Relationship Between Measured Binding Strength and Standard Error of the RBA
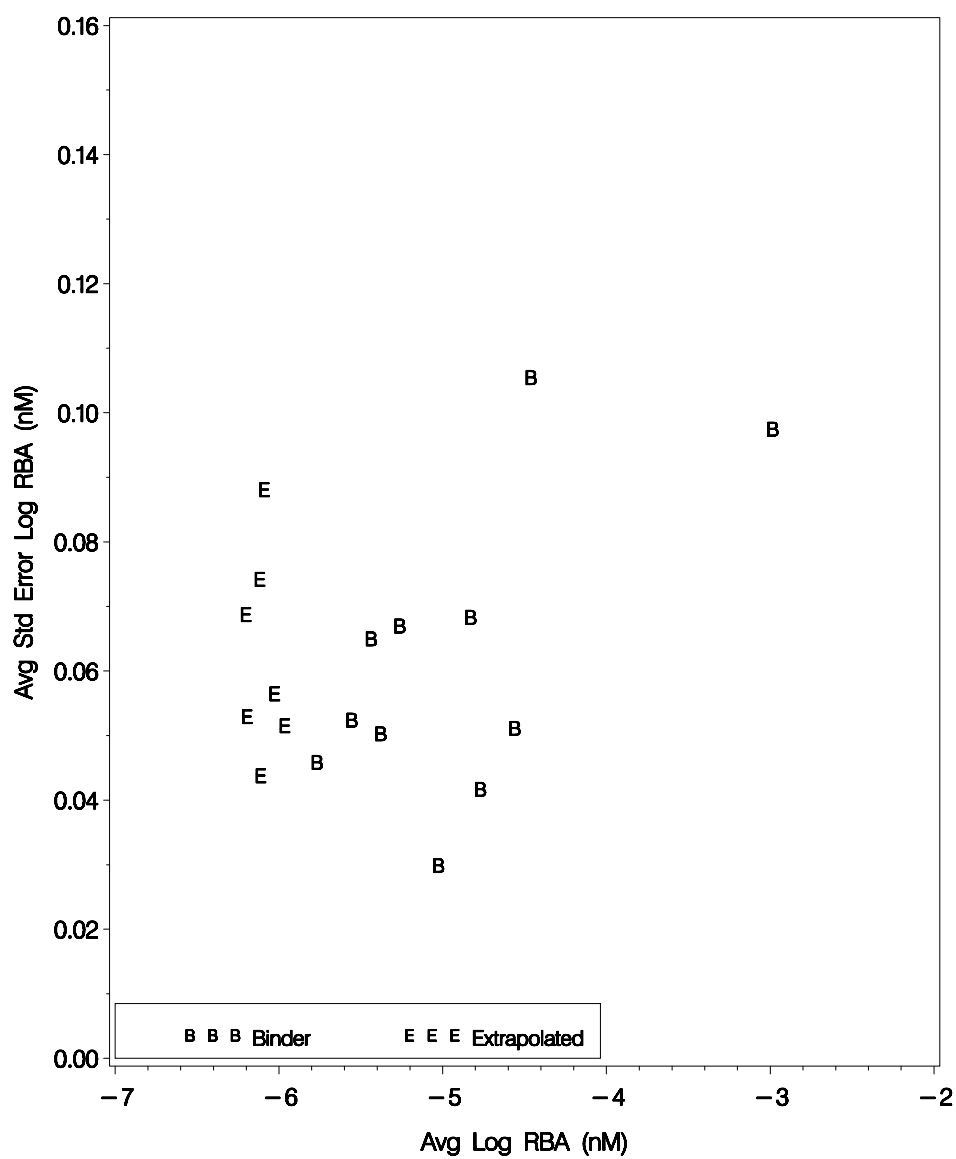'Random 200' Chemicals Restricted to Binders or Extrapolated

Figure A-1

Relationship Between Measured Binding Strength and Standard Error of the RBA
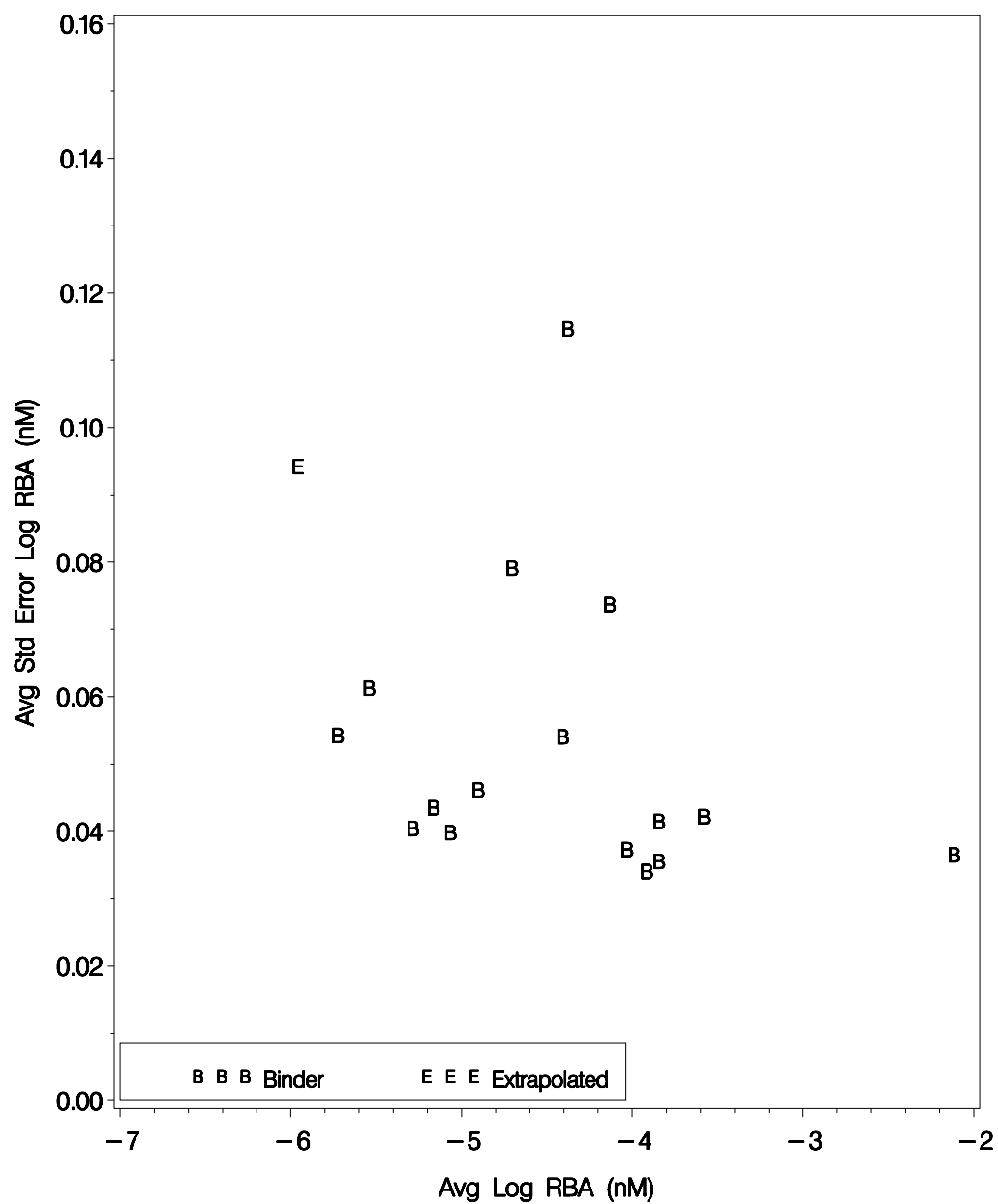Lab Analyzed Model A Chemicals Restricted to Binders or Extrapolated

Figure A-2

Relationship Between Measured Binding Strength and Standard Error of the RBA
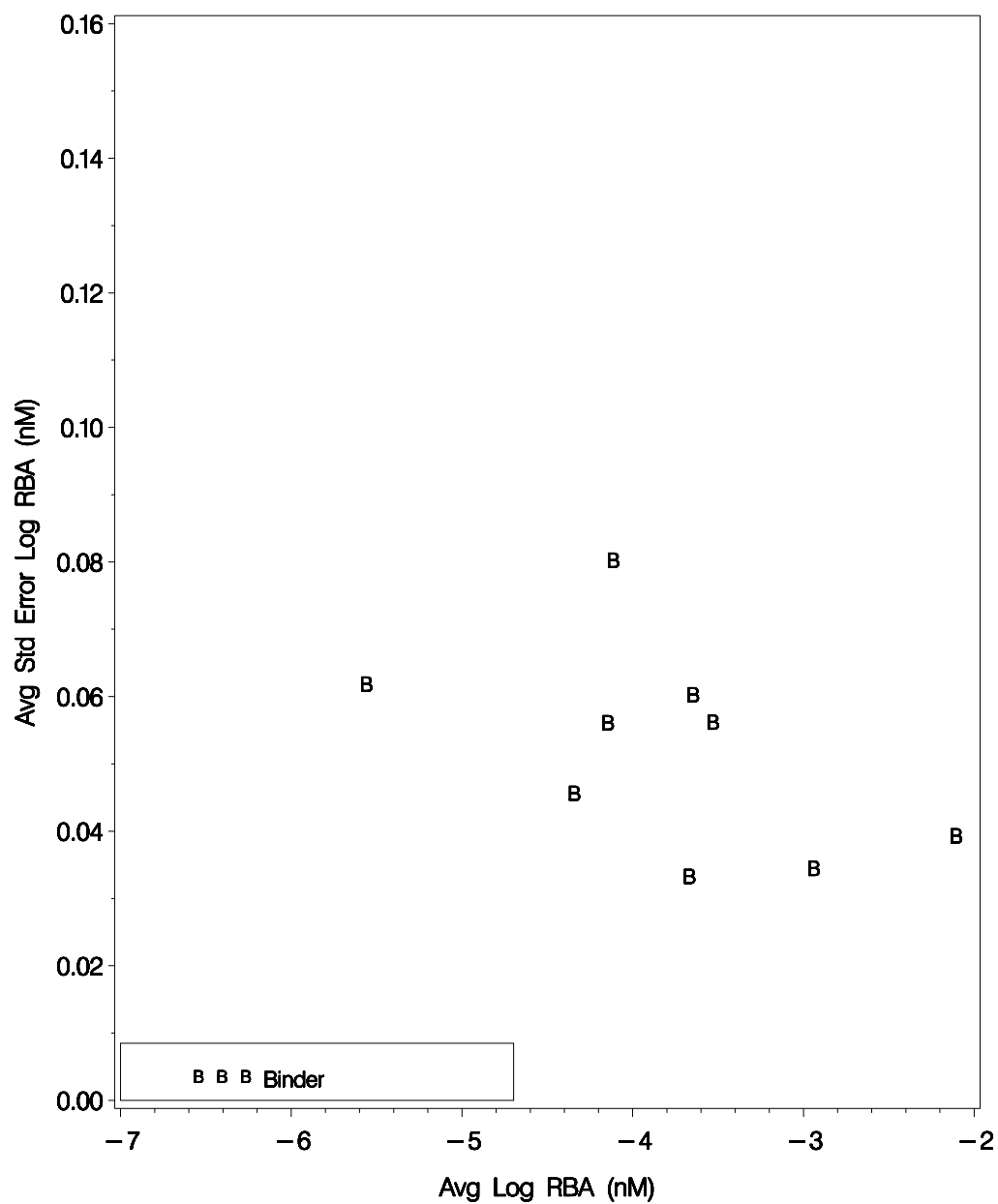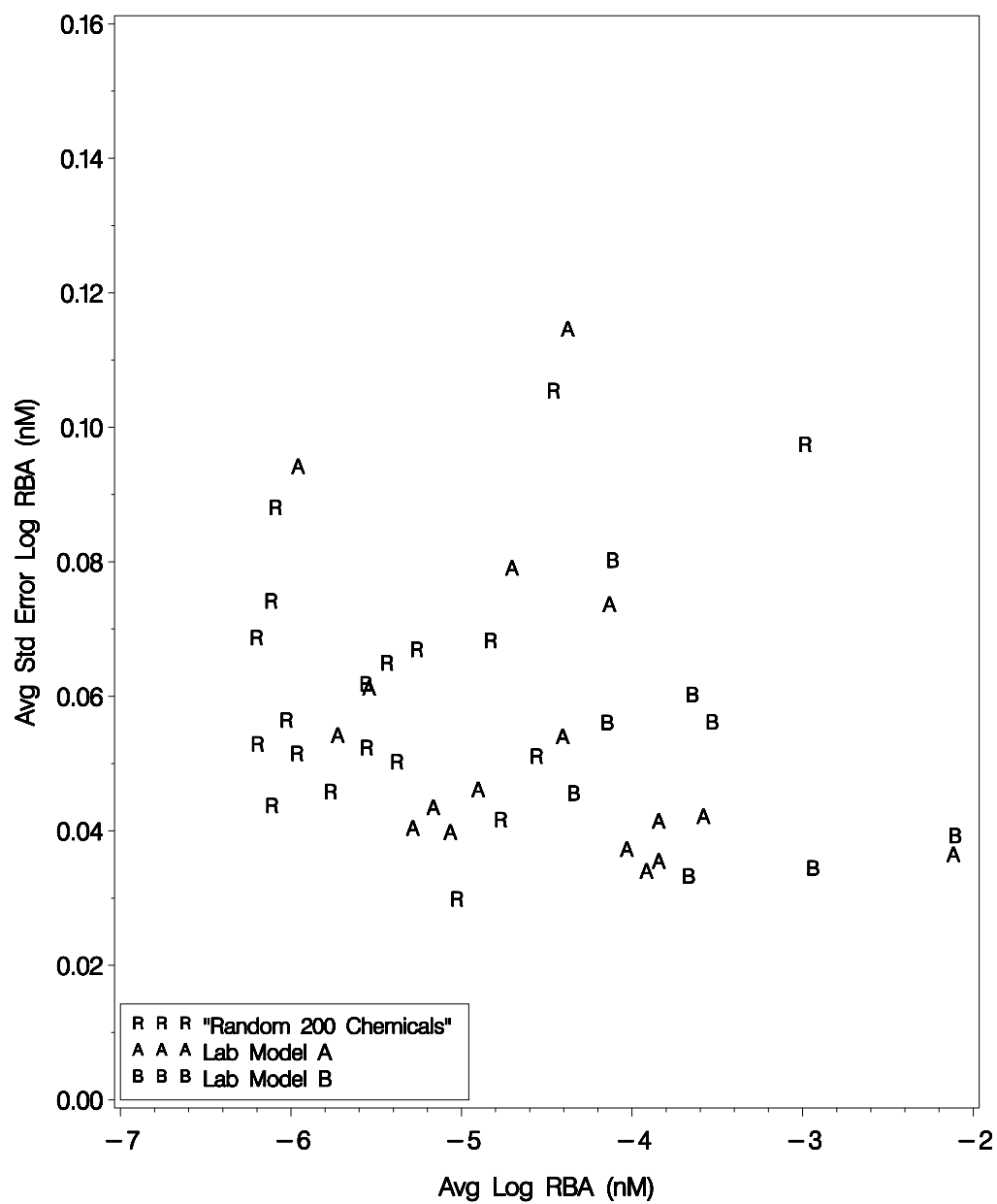Lab Analyzed Model B Chemicals Restricted to Binders or Extrapolated

Figure A-3

Relationship Between Measured Binding Strength and Standard Error of the RBA
Combined Chemical Data Restricted to Binders or Extrapolated

Figure A-4

APPENDIX B

CAS NUMBERS INCLUDED IN

"Random 200" Sample
"50 Model A" Positive Prediction Sample
"50 Model B" Positive Prediction Sample

[deleted from this copy – jpk, 8-2-02]